Evidence-Based Elections

P.B. Stark and D.A. Wagner

***Abstract*—We propose an alternative to current requirements for certifying voting equipment and conducting elections. We argue that elections should be structured to provide convincing affirmative evidence that the reported outcomes actually reflect how people voted. This can be accomplished with a combination of software-independent voting systems, compliance audits, and risk-limiting audits. Together, these yield a resilient canvass framework: a fault-tolerant approach to conducting elections that gives strong evidence that the reported outcome is correct or reports that the evidence is not convincing. We argue that, if evidence-based elections are adopted, certification and testing of voting equipment can be relaxed, saving money and time and reducing barriers to innovation in voting systems—and election integrity will benefit. We conclude that there should be more regulation of the evidence trail and less regulation of equipment, and that compliance audits and risk-limiting audits should be required.**

***Keywords*-elections, software-independent voting system, risk-limiting audit, resilient canvass framework EDICS SEC-INTE, APP-CRIM, APP-INTE, APP-OTHE.**

I. Introduction

IDEALLY, what should an election do? Certainly, an election should find out who won, but we believe it also should produce convincing evidence that it found the real winners—or report that it cannot. This is not automatic; it requires thoughtful design of voting equipment, carefully planned and implemented voting and vote counting processes, and rigorous post-election auditing.

The systems and processes currently deployed in the US often fail to meet this goal, due to shortcomings of the equipment, gaps in the processes, and failures to audit effectively. The first essential requirement is voting equipment that produces a trustworthy audit trail that can be used to confirm that the votes were recorded and tabulated correctly. Given the present state of technology, this means the voting system must produce a tangible, physical record of the vote that can be checked by the voter for accuracy and retained for auditing purposes: typically, a *voter-verifiable paper record*. While approximately 75% of US voters currently vote on equipment that produces a voter-verifiable paper record of the vote, about 25% vote on paperless electronic voting machines that do not produce such a record [1].

Because paperless electronic voting machines rely upon complex software and hardware, and because there is no feasible way to ensure that the voting software is free of bugs or that the hardware is executing the proper software, there is no guarantee that electronic voting machines record the voter's votes accurately. And, because paperless voting machines preserve only an electronic record of the vote that cannot be directly observed by voters, there is no way to produce convincing evidence that the electronic record accurately reflects the voters' intent. Internet voting shares the shortcomings of paperless electronic voting machines, and has additional vulnerabilities.

Numerous failures of electronic voting equipment have been documented. Paperless voting machines in Carteret County, North Carolina irretrievably lost 4,400 votes; other machines in Mecklenburg, North Carolina recorded 3,955 more votes than the number of people who voted; in Bernalillo County, New Mexico, machines recorded 2,700 more votes than voters; in Mahoning County, Ohio, some machines reported a negative total vote count; and in Fairfax, Virginia, county officials found that for every hundred or so votes cast for one candidate, the electronic voting machines subtracted one vote for her [2]. In short, when elections are conducted on paperless voting machines, there is no way to produce convincing evidence that the right candidate won.

Voter-verifiable paper records are important, but they are not a panacea. If these records are not examined after the election, then their value is eliminated. For instance, in 13 states, a voter-verifiable paper record of every vote is produced but not audited after the election, with the consequence that observers and candidates do not receive convincing evidence that the election outcome reflects the will of the voters [1].

A natural reaction to growing concern about electronic voting machines is to call for a stricter certification process. Perhaps counter-intuitively, we argue in this paper that stricter certification processes may not achieve the desired goals and may have unintended consequences that render the approach counter-productive.

Currently, states generally require jurisdictions to use voting systems that have been tested against a federal or state standard. Historically, these standards were written primarily for electro-mechanical voting systems, such as lever machines, paper ballots, and punchcard voting systems. However, the trend over the past decade or two has been towards more sophisticated, complex technology and much greater reliance upon complex software—trends that voting standards and the testing process have been slow to react to. Perhaps as a result,

PBS: Department of Statistics, University of California, Berkeley CA, 94720-3860, USA. e-mail: stark AT stat DOT berkeley DOT edu. DAW: Department of Electrical Engineering and Computer Science, University of California, Berkeley CA, 94720-1776, USA. e-mail: daw AT cs DOT berkeley DOT edu.

This work was partially supported by National Science Foundation grant CNS-0524745. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We thank Ben Adida, Jennie Bretschneider, Joshua M. Franklin, Lynn Garland, David Jefferson, Mark Lindeman, Neal McBurnett, Dan McCrae, Ronald L. Rivest, and the anonymous reviewers for helpful comments on earlier drafts.

Manuscript received 9 January 2012; revised 18 March 2012. First published ? October 2012.

independent studies of deployed voting software have found major design defects and security vulnerabilities. For instance, one study commissioned by the state of Ohio found pervasive security flaws of sufficient severity that the researchers concluded the equipment would need "fundamental and broad reengineering" to support the goal of guaranteeing trustworthy elections. Flaws included default passwords, unprotected software upgrade functionality, and defects that would allow a single individual—with no insider access—to introduce a malicious virus that spreads from voting machine to machine and silently changes votes [3]. Over a dozen independent studies have evaluated the security of fielded, certified voting systems, and every one found new, previously unknown security flaws that certification had not uncovered. One flaw had been discovered independently at least three times by different security experts over a period of nine years (in 1997, 2003, and 2006), but was never caught by certification testing [4]. (Since then, the US Election Assistance Commission began to administer federal certification and made a number of changes to the process.)

Accordingly, many computer scientists have called for voting standards to be tightened and for the certification process to be made more rigorous [5], [6], [7], [8]. The US Election Assistance Commission (EAC) has taken modest steps in this direction. For instance, the certification process is more stringent than ever, and the EAC is developing stricter standards for next-generation voting technology. However, a consequence of these changes is that the cost of certifying a voting system has risen dramatically. While the time and costs of certification depend on many factors, generally speaking, federal certification testing can cost manufacturers about $2–4 million to certify a new voting system and can take up to two years [9]. Many fear that the cost will rise even higher if future standards are adopted [10], [11], [12].

At the same time, the voting industry has consolidated: Two companies now dominate the market, and the expense of certification testing makes it exceptionally difficult for new players to enter the market. Moreover, even these more exacting certification processes do not ensure that electronic vote records will accurately reflect the intent of the voter nor that electronic vote records provide convincing evidence of the election's accuracy. One could reasonably ask whether the costs of certification outweigh the benefits, let alone whether it is wise to make certification even more onerous.

Certification tries to prevent problems before they occur. The complexity of voting technology, software, and procedures makes it impossible even to enumerate all the features of the system and circumstances of use that might cause problems, much less to prevent all problems. The difficulty is exacerbated by the fact that certification tests new equipment under a limited range of conditions. How systems are maintained, configured, and deployed, and the particular procedures a jurisdiction follows, obviously affect reliability and accuracy. We argue below that putting more emphasis on detecting and correcting errors and less on certification is both more economical and more effective at ensuring election integrity.

II. Evidence-based elections

We sketch how evidence-based elections might be conducted, using several technical tools. The basic approach we advocate is simple:

$$\text{evidence} = \text{auditability} + \text{auditing}.$$

In other words, the voting equipment must be auditable: It must produce a trustworthy audit trail that can be used to check the accuracy of the election. In addition, election processes must incorporate auditing as part of the routine electoral process, to produce and check evidence that the election found the right winners.

The technical name for the approach we advocate is a *resilient canvass framework* [13]. This can be achieved by a *strongly software-independent voting system* [14], [15], which provides auditability by generating an audit trail that can be used to find the actual winners, and two kinds of routine post-election audits: a *compliance audit* and a *risk-limiting audit* [16], [17], [18], [19], [20], [21]. The compliance audit checks that the audit trail is sufficiently complete and accurate to tell who won. The risk-limiting audit checks the audit trail statistically to determine whether the vote tabulation system found the correct winners, and, with high probability, corrects the outcome if the vote tabulation system was wrong.

All three components are crucial. The risk-limiting audit relies on the integrity of the audit trail, which was created by the software-independent voting system (the voters themselves, in the case of paper ballots) and checked for integrity by the compliance audit. We now sketch the three ingredients in greater detail.

A. Strongly software-independent voting systems

A voting system is strongly software-independent [14], [15] if an undetected error or change to its software cannot produce an undetectable change in the outcome, and we can find the correct outcome without re-running the election. Strong software-independence does not mean the voting system has no software; rather, it means that even if its software has a flaw that causes it to give the wrong outcome, the overall system still produces "breadcrumbs" (an audit trail) from which we can find the true outcome, despite any flaw in the software. Systems that produce voter-verifiable paper records (for instance, voter marked paper ballots) as an audit trail are strongly software-independent, provided the integrity of that audit trail is maintained, because the audit trail can be used to determine who really won. Currently, the only systems that can confer software independence are end-to-end cryptographic systems and systems with voter-verifiable paper records (VVPRs). Voter marked paper ballots are the best form of VVPR. Thermal paper print of the kind typically generated by direct-recording electronic voting machines (DREs) is particularly fragile and subject to spoilage.

B. Compliance audits

The voting equipment is responsible for generating audit records (e.g., VVPRs) that are trustworthy at the time they were generated. However, this is not enough. We must also ensure that they remain trustworthy at the time they are used in the audit.

Normally, we rely upon other election processes to protect the audit trail. These processes include two-person chain-of-custody rules, tamper-evident seals, ballot accounting, and other procedural methods. The purpose of a compliance audit is to check whether these steps were properly followed. For instance, if the voting system relies on VVPRs to achieve strong software independence, the compliance audit is responsible for generating affirmative evidence that procedures have been followed, so that the VVPRs will be a sufficiently accurate and complete record of voter intent to determine who won.

Ultimately, the purpose of a compliance audit is to generate convincing affirmative evidence that a full hand count of the audit trail would correctly reflect the outcome of the election, as the votes were cast. If the election passes the compliance audit, we may proceed to the risk-limiting audit. However, if the compliance audit cannot confirm that the audit trail is trustworthy, then the overall canvass process must be deemed a failure: It did not produce convincing evidence that the election was decided correctly.

Compliance audits are not a silver bullet. Elections will still rely upon people, processes, and procedures. Compliance audits do not give ironclad proof that all procedures were faithfully followed, so the human element—poll workers and others people involved in running the election—remains an important part of election administration. Still, compliance audits can detect systemic procedural failures and help measure whether the audit trail is trustworthy.

A compliance audit might include the following steps:

- Poll book accounting: Compare the number of voters who signed in to vote in each precinct to the number of ballots cast. Investigate and document discrepancies.
- Ballot accounting: Check that the number of blank ballots distributed to each precinct matches the number of blank/spoiled ballots returned plus the number of ballots cast in that precinct. Check that the number of ballots voted by mail is not greater than the number mailed to voters.
- Chain of custody checks: Examine chain of custody signature logs. Confirm that cast ballots were never left unattended without video surveillance nor in the sole possession of any one individual.
- Security checks: Check seals, seal logs, and surveillance tapes. Voted ballots should be protected with tamper-evident seals for transport to and storage at election headquarters. Check that seal procedures have been properly followed.

  For instance, the county might include a uniquely numbered seal in each precinct's supplies and record the numbers of the seals sent to each precinct. After the polls close, poll workers could apply the seal to the box of voted ballots before transport, photograph the applied seal, and upload the photograph to the county. When the box is received at a county collection center, county workers could check the seals and seal numbers, and compare the seals to the photographs. If ballots need to be inspected as part of the audit, seal integrity can be checked again. If seals are inadvertently broken, replaced, or other anomalies occur, these events should be logged. The compliance audit should check that these protocols have been followed.
- Event log inspection: Check voting system logs, poll worker problem logs, and other materials.

Many election offices may already conduct some or all of these checks.

To facilitate transparency and give the public convincing evidence that the election found the correct outcome, the results of the compliance audit should be made available to observers, candidates, and members of the public upon request. Because none of this information can compromise the secrecy of the ballot, it can be published without endangering voter privacy. In addition, election officials should review the results of the compliance audit after each election and use them to continuously improve election procedures.

If the compliance audit discovers problems or gaps in the evidence, the next step is to determine which audit records are trustworthy and which might not be. This information can be fed into the risk-limiting audit. If the number of questionable or missing audit records is small enough that they cannot alter the outcome of the contest, a risk-limiting audit may still be able to provide strong evidence of who won, while treating the questionable or missing records in the most pessimistic way.

C. *Risk-limiting audits*

Risk-limiting audits involve quantifying the evidence that the reported electoral outcome is correct, by manually inspecting parts of the audit trail—which the compliance audit has already confirmed is sufficiently accurate to determine who won. Typically, a risk-limiting audit examines more and more of the audit trail until the manual inspection generates convincing evidence that a full hand count of the audit trail would show the same outcome that the vote tabulation system reported, or until it has examined the entire audit trail. If looking at a portion of the audit trail gives sufficiently strong evidence that a full hand count would confirm the outcome, the risk-limiting audit can stop: We have confidence that the reported winners of the election are the real winners. If the audit ends up examining all ballots, effectively performing a 100% hand count, we can certify with confidence that the actual outcome is the outcome that the full hand count shows.

Of course, audits are not guaranteed to find all errors. Risk-limiting audits have a large chance of correcting the outcome if the outcome is wrong, but do not promise to detect or correct errors that, aggregated, could not change the outcome. The confidence level—100% minus the risk limit—is a guaranteed lower bound on the probability that the audit will correct the outcome if the outcome is wrong, no matter why the outcome is wrong. Statistical methods can be used to determine when the audit has achieved any desired confidence level, e.g., 90% (10% risk). In general, higher confidence levels (lower risk limits) require inspecting more records, to obtain stronger evidence.

Other work in this volume gives simple methods for two basic types of risk-limiting audits [21]. One type (comparison audits) works by randomly selecting groups of ballots and comparing the vote subtotal from the vote tabulation system for each group against a hand count of the ballots in that group. Comparison audits are more efficient when the groups contain fewer ballots, and most efficient when each group consists of a single ballot ("ballot-level audits"). Ballot-level comparison audits also increase transparency: It is much easier for auditors and observers to check whether a particular ballot shows a vote for a particular candidate than to verify that the votes in a group of dozens or hundreds of ballots were tabulated correctly.

Comparison audits place demands on vote tabulation systems. The system must be able to report, in a useful format, the vote subtotals for each group of ballots. There is some flexibility in choosing how ballots are grouped (for instance, one can group ballots by precinct, separating those cast at the polling place from those cast by mail), but the groups must be defined so that all the ballots in a single group can be readily identified and retrieved without compromising voter privacy.

As discussed below, current vote tabulation systems make auditing unnecessarily difficult and expensive because they do not report vote subtotals for groups in a directly useful format and because they cannot report results for arbitrarily small groups. In addition, most current systems cannot report their interpretations of individual ballots ("cast vote records"), which would be required for ballot-level comparison audits.

Risk-limiting audits have been field tested in a number of jurisdictions, on contests of various types and sizes. In California, there have been audits in the counties of Alameda, Humboldt, Marin, Merced, Orange, San Luis Obispo, Santa Cruz, Stanislaus, Ventura, and Yolo. There have also been risk-limiting audits in Boulder County, Colorado, and Cuyahoga County, Ohio. California and Colorado received grants from the EAC to conduct pilot risk-limiting audits. Colorado Revised Statutes 1-7-515 requires risk-limiting audits by 2014. California AB 2023 required a pilot of risk-limiting audits in 2011. Risk-limiting audits can handle many forms of contests, including multi-candidate contests, vote-for-n contests, and measures requiring a super-majority. Multiple contests of different types can be audited simultaneously with a single sample. Risk-limiting audits of ranked-choice voting and instant-runoff voting are possible in principle, but are opaque and computationally demanding [22], [23]. Pilots of risk-limiting audits have been applied to elections from 200 to 121,000 ballots; they have required inspecting from 16 to 7,000 ballots. This experience suggests that risk-limiting audits are practical and inexpensive compared to other election costs [17], [18], [24].

D. Resilient canvass frameworks

Ideally, the overall election and canvass process should correct its own errors before announcing results, or report that it cannot guarantee that it corrected its errors (for instance, because the compliance audit finds gaps in the chain of custody for many ballots or finds that there are far fewer cast ballots than poll book signatures). The three-part framework described above is designed to achieve exactly this goal. If the system fails the compliance audit, the process reports that it cannot guarantee the correctness of the outcome; otherwise, a risk-limiting audit is conducted to check the election outcome and—with high probability—to correct the outcome if it is wrong.

This notion has been formalized in the literature under the name *resilient canvass framework* [13]. A canvass framework (including humans, hardware, software, and procedures) is resilient if it has a known minimum chance of giving the correct election outcome—when it gives an outcome. Prior work shows that our proposed three-part framework—the combination of strongly software-independent voting systems, compliance audits, and risk-limiting audits—achieves this goal [13].

Resilience combines two kinds of evidence, qualitative and quantitative. The compliance audit generates qualitative evidence, much like legal evidence. For instance, we might ask the compliance audit whether it is "beyond reasonable doubt" that the audit trail reflects the true outcome. The kind of evidence the compliance audit generates resists objective quantification.

The compliance audit can help us decide how much we can rely on the audit trail. For instance, if the compliance audit finds that a single innocent mistake could alter the outcome without leaving a trace in the audit trail—as is the case with paperless voting machines—the evidence that the outcome is correct is not persuasive. If the compliance audit discovers that the number of ballots missing or left unattended could account for the margin several times over, the evidence that the outcome is correct is not persuasive. In contrast, if it would take dozens of bad actors, both insiders and outsiders, conspiring undetected, to alter the reported outcome and to alter the audit trail so that it reflects that wrong outcome, it is reasonable to conclude that a full examination of the audit trail would reveal the true outcome.

The risk-limiting audit generates quantitative evidence. If the audit trail reflects the true outcome, it is straightforward in principle to find the minimum probability that the risk-limiting audit would correct the outcome if the outcome is wrong: The auditors themselves control that probability by how they draw the sample and how they decide when to stop the audit. There are now many rigorous methods for risk-limiting audits with varying requirements on voting systems, varying levels of complexity, and varying levels of efficiency (e.g., [16], [18], [19], [20], [21]).

III. Certification

A. The costs and shortcomings of voting-system certification

Consider two related questions about the tabulation of votes in elections:

1) Under laboratory conditions, can the vote tabulation system—as delivered from the manufacturer—count votes with a specified level of accuracy?
2) As maintained, deployed, and used in the current election, did the vote tabulation system find the true winners?

We believe that question 2 is the more important and that question 1 is interesting primarily insofar as it bears on question 2. Systems that can function accurately under ideal conditions might not have functioned well as maintained, programmed, and deployed. While there is undeniably value to weeding out flawed voting systems before they can be used in an election, ultimately evidence that the voting system actually found the true winners matters more than evidence that, under laboratory conditions, it can count votes well. Certification of voting systems addresses question 1. Post-election audits address question 2.

Current certified vote tabulation systems make risk-limiting audits unnecessarily difficult, expensive, and opaque. The most efficient and transparent risk-limiting audits involve comparing the system's interpretation of individual ballots to a human interpretation of the same ballots. However, no federally certified vote tabulation system reports how it interpreted individual ballots, making it difficult to conduct ballot-level audits. If a jurisdiction uses a certified vote tabulation system, it will cost more to use it as a component of a resilient canvass framework because auditing will be more expensive, and the audit may be less convincing to observers than if the jurisdiction had used an uncertified system that does support efficient, transparent auditing.

Currently certified vote tabulation systems cannot report vote subtotals (in a useful, machine-readable format) for the sorts of groups of ballots needed to support efficient risk-limiting audits. For instance, they may be able to report vote subtotals by precinct, but since vote-by-mail ballots may not be sorted or retrievable by precinct, this is not sufficient; and tabulation systems often cannot report vote subtotals for the groups that vote-by-mail ballots are scanned in. Moreover, when these systems can report vote subtotals, the reports are not designed for machine processing and consequently a great deal of hand editing is necessary to use them in an audit. Certified vote tabulation systems cannot report results for arbitrarily small groups of ballots; typically, they can generate reports for precincts subdivided by mode of voting (in person or by mail), but not smaller groups. To produce machine-readable reports for smaller groups—such as ballots scanned as a batch—might require only a few lines of SQL, but to add those lines to current systems would require re-certification. As a result, the current requirement for certification impacts election integrity by making risk-limiting audits less efficient, less transparent, and more expensive.

Since currently certified systems make risk-limiting audits expensive, inefficient, and untransparent, perhaps it would be better to allow the use of uncertified vote tabulation systems (or systems certified to a less stringent standard) that are easier to audit, provided electoral outcomes *are* then checked with risk-limiting audits. An uncertified tabulation system that supports ballot-level auditing could increase both efficiency (fewer ballots need to be examined manually to attain the same level of confidence) and transparency (it is easier to observe whether a particular ballot shows a vote for a particular candidate than it is to observe whether a large number of ballots were tallied correctly).

In general, the current certification regime has drawbacks beyond its effect on auditing. The cost of certification adds to the cost of voting systems, at a time when election officials have very tight budgets. The high cost of new voting systems locks jurisdictions into using old equipment long past its obsolescence. The cost of certification is a barrier to entry for competitors, especially for small players, and tilts the playing field towards large vendors. This encourages consolidation, reducing the number of options available to election officials. The cost and time for re-certification make it harder to patch bugs in certified systems or add additional features. Also, because entire systems currently must be certified as a whole (there is no provision to certify a single component on its own), it is difficult to update individual commercial, off-the-shelf (COTS) components of the voting system, such as digital scanners, which would be commodity items but for the certification requirement. The consequence is that certification may inhibit innovation and efficiency in voting technology and increase up-front costs and maintenance costs.

B. More focused roles for certification, legislation, regulation, and advice

We suggest that certification should focus on ensuring that a reliable audit trail is generated, rather than trying to ensure that votes are tabulated accurately—a role certification cannot fulfill. Procedures required by laws and regulations can then ensure that the audit trail remains reliable throughout the canvass; that the reliability is confirmed by a compliance audit; and that the audit trail is in fact audited to generate convincing evidence that the reported outcome is correct—or to correct it.

Certification may be the best tool to address usability and accessibility, including ballot design and ballot presentation, to ensure that all voters can easily record their preferences. It might also be the best way to address some conditions required to maintain voter privacy, for instance, to guarantee that electronic voting machines do not inadvertently record information that could be used to link a particular voter to a particular ballot. Certification of vote tabulation accuracy in the laboratory is less important, because the accuracy of tabulation *in the current election* can be checked after the fact. In contrast, presenting voters with a confusing ballot cannot be rectified without re-running the election. In addition, since many aspects of usability cannot be measured in the field directly during a real election without violating voter privacy, usability can only be measured in advance in the laboratory.

Certification testing may also be the best tool to protect against failures that, if they were to occur, could not be recovered from at any acceptable cost. If a technology failure can prevent voters from marking, verifying, or casting their ballots, then certification should include reliability testing to minimize the frequency of these failures during elections. (For instance, DREs with VVPRs would require careful testing to ensure that the voting machines are reliable enough that voters will be able to vote successfully. In contrast, because voters can continue to mark and cast paper ballots even in the event of a technology failure, optical scanners might not need reliability testing.) As another example, because a failure of a vote tabulation system that prevents election officials from reporting unofficial results on election night can cause bad publicity and loss of confidence, election officials might choose to include some level of reliability testing to check that the voting system will be able to tabulate votes on election night.

Further work will be required to define the appropriate role of standards and testing in detail. Based upon these principles, our framework might allow eliminating source code review, restrictions on coding conventions, and similar certification testing requirements. Certain elements proposed for next-generation voting standards might also be eliminated from the testing process, e.g., open-ended vulnerability testing, logic verification, and testing of setup inspection, cryptographic hardware, access control functionality, and other security features.

For voting systems such as optically scanned voter marked paper ballots, where voters can continue to vote even if the hardware or software fails, additional elements could be eliminated: volume testing, environmental testing and other forms of reliability testing, restrictions on the use of COTS components, and so on. However, for DREs and other systems whose failure would prevent voters from marking or casting their ballot on election day, reliability tests would still be needed.

Alternatively, vendors could self-assess or self-attest their compliance with such requirements, and regulators could monitor the performance of voting systems in the field. Vendors and test labs have found several of these elements—particularly source code review, open-ended vulnerability testing, and volume testing—to be especially expensive [11], [12], so eliminating them from the testing regime could significantly reduce the cost of certification testing.

Other elements of the current and proposed testing regime would still be needed, including testing usability and accessibility and testing the core functional requirements for counting and tabulation of votes and vote privacy properties. Test labs would also need to confirm that the voting system can meet the software independence requirement and can support efficient risk-limiting audits and compliance audits.

Election officials could decide whether the certification process retains testing of certain functional requirements, such as pre-election programming, ballot design, support for logic and accuracy tests, or whether to rely upon other mechanisms for quality assurance, such as procurement processes, market competition, and in-the-field monitoring. While testing these requirements is not necessary for auditing and not required by our framework, election officials nonetheless may find such tests useful for purchase decisions.

In our view, laws should enunciate the basic requirements as principles: Use a system that generates a reliable audit trail (use a strongly software-independent system); check that the audit trail remained reliable throughout the canvass (perform a compliance audit); and examine the audit trail in a way that has a large, pre-specified chance of correcting the outcome before it is official if the outcome was wrong (perform a risk-limiting audit at a specific risk limit).

State regulations should then provide “safe harbor” methods and procedures. For instance, state regulations might include a minimal checklist for compliance audits and guidelines for publishing election data, and might spell out a simple, step-by-step procedure for conducting a risk-limiting audit (see, e.g., [25]). State regulations should also specify how to evaluate whether other proposed procedures meet the legal standard, for jurisdictions that want to use procedures more closely tailored to their individual needs.

Federal or state agencies might act as clearinghouses for information about voting systems, including mandatory reports of failures of accuracy or reliability. They might still test equipment and software, to be able to endorse some systems as having performed reliably, preferably after field testing, not just laboratory testing. Such endorsements may save jurisdictions time, trouble, and money, while—as we have argued—certification tends to increase local expenditures without necessarily producing better evidence that election outcomes are correct.

Mandating compliance and risk-limiting audits would give local elections officials strong incentives to find the most accurate, reliable, and economical solution within their constraints: The more accurate the initial machine count, the less hand counting the audit would require. And because development, acquisition, and maintenance costs would be lower and more COTS components could be used, the entire system would provide stronger evidence at lower cost to society, and maintain the agility to adopt newer and better solutions quickly. That said, our proposal is a significant change from the status quo: To understand the cost and regulatory implications will require more study.

IV. Discussion

We believe that election law should require local election officials to give convincing evidence that election outcomes are correct. The call for “convincing evidence” raises the question, “convincing to whom?” We do not probe this question deeply here, but we think that the compliance audit should generate evidence that is convincing to “a reasonable person,” a common legal standard. That is, after the compliance audit it should be “beyond reasonable doubt” that the audit trail is adequately intact to determine the true winners; otherwise, the election has failed the compliance audit. The risk-limiting audit should provide strong, quantitative, statistical evidence: evidence that would convince anyone who understands the basis of the calculations or at least believes the theory behind the calculations. The required strength of evidence—the risk limit or confidence level—can be set by legislation.

Providing convincing evidence does not require radical transparency of all election and canvass processes, but it does require a good audit trail, affirmative evidence that the audit trail is reliable, and adequate scrutiny of the audit trail to confirm that the votes were tabulated accurately enough to determine the true winners. To that end, we believe that there should be more focus on regulating procedures, especially the curation of the audit trail, and less focus on certifying tabulation equipment, in part because certification can never guarantee that votes are tabulated accurately in practice: How the system is maintained and deployed and how the data are handled are crucial. If the system generates a reliable audit trail and that trail is curated well, a risk-limiting audit can check whether—in the current election, given how the equipment was maintained and used—the votes in fact were tabulated accurately enough to determine the correct winners, and can guarantee a large chance of correcting the results if not.

Currently, certification does not serve the interests of the public or local elections officials as well as one might hope. It erects barriers to competition, increases acquisition and maintenance costs, slows innovation, and makes risk-limiting audits harder, slower, more expensive, and less transparent. And risk-limiting audits provide more direct evidence that outcomes are correct than certification can. Requiring local election officials to conduct compliance and risk-limiting audits rather than requiring them to use certified equipment would give them an additional incentive to use the most accurate (and most easily audited) tabulation technology available, because more accurate initial counts require less hand counting during the audit, reducing labor costs and allowing the canvass to finish sooner.

We believe that election integrity would be served best by laws and regulations that put incentives in the right place, and that focus on evidence rather than equipment and software. Focusing on evidence entails more attention to creating, curating, and auditing the integrity of the audit trail, including requiring and and checking seals and surveillance, chain of custody logs, ballot accounting, and other safeguards. Using voting systems that are designed to support efficient auditing—whether those systems have been certified or not—can substantially reduce the cost of collecting convincing evidence that the official results are correct.

References

[1] http://verifiedvoting.org, retrieved 8 January 2012.

[2] Common Cause and VotersUnite!, "A Master List of 70+ Voting Machine Failures and Miscounts by State," retrieved 17 March 2012. [Online]. Available: http://www.commoncause.org/atf/cf/%7Bfb3c17e2-cdd1-4df6-92be-bd4429893665%7D/MASTERLISTOFMACHINEFAILURES.PDF

[3] P. McDaniel, M. Blaze, G. Vigna, and et al., "EVEREST: Evaluation and Validation of Election-Related Equipment, Standards and Testing," Dec 2007, retrieved 17 March 2012. [Online]. Available: http://www.sos.state.oh.us/SOS/upload/everest/14-AcademicFinalEVERESTReport.pdf

[4] D. Wagner, "Testimony before U.S. House of Representatives at joint hearing of the Committee on Science and Committee on House Administration," Jul 2006, retrieved 17 March 2012. [Online]. Available: http://www.cs.berkeley.edu/~daw/papers/testimony-house06.pdf

[5] ACCURATE, "Public Comment on the Voluntary Voting System Guidelines, Version 1.1," Sep 2009, retrieved 17 March 2012. [Online]. Available: http://accurate-voting.org/wp-content/uploads/2009/09/ACCURATE-vvsgv11-final.pdf

[6] ——, "Public Comment on the 2005 Voluntary Voting System Guidelines," Sep 2005, retrieved 17 March 2012. [Online]. Available: http://accurate-voting.org/accurate/docs/2005_vvsg_comment.pdf

[7] D. Mulligan and J. L. Hall, "Preliminary Analysis Of E-Voting Problems Highlights Need For Heightened Standards And Testing," Dec 2004, retrieved 17 March 2012. [Online]. Available: http://josephhall.org/papers/NRC-CSTB_mulligan-hall_200412.pdf

[8] E. W. Felten, "Testimony, United States House of Representatives, Committee on House Administration Hearing on Electronic Voting Machines: Verification, Security, and Paper Trails," Sep 2006, retrieved 17 March 2012. [Online]. Available: http://usacm.acm.org/images/documents/felten_testimony.pdf

[9] US Dept. Justice v. ES&S, Mar 2010, US District Court for the District of Columbia, case no. 1:10-cv-00380. [Online]. Available: http://www.justice.gov/atr/cases/f256200/256275.htm

[10] D. Beirne, "Written Remarks, submitted to the United States Election Assistance Commission Interdisciplinary Roundtable Discussion on the Proposed Voluntary Voting System Guidelines," May 2008, retrieved 17 March 2012. [Online]. Available: http://archives.eac.gov/News/docs/080505roundtableremarksfinal_davidbeirne/attachment_download/file

[11] H. S. Berger, "Testimony Concerning the TGDC 2007 Draft Revision of the Voluntary Voting System Guidelines," Mar 2008, retrieved 17 March 2012. [Online]. Available: http://archives.eac.gov/News/meetings/testimony-draft-vvsg-2007-080318/attachment_download/file

[12] "United States Election Assistance Commission Public Meeting Voting Systems Manufacturer Roundtable Discussion: Verbatim Transcript," Feb 2008, retrieved 17 March 2012. [Online]. Available: http://archives.eac.gov/News/docs/02-29-08-transcript-public-meeting/attachment_download/file

[13] J. Benaloh, D. Jones, E. Lazarus, M. Lindeman, and P. Stark, "SOBA: Secrecy-preserving observable ballot-level audits," in *Proceedings of the 2011 Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (EVT/WOTE '11)*. USENIX, 2011. [Online]. Available: http://statistics.berkeley.edu/~stark/Preprints/soba11.pdf

[14] R. Rivest and J. Wack, "On the notion of "software independence" in voting systems (draft version of july 28, 2006)," Information Technology Laboratory, National Institute of Standards and Technology, Tech. Rep., 2006, http://vote.nist.gov/SI-in-voting.pdf Retrieved 17 March 2012.

[15] R. Rivest, "On the notion of 'software independence' in voting systems," *Phil. Trans. R. Soc. A*, vol. 366, no. 1881, pp. 3759–3767, October 2008.

[16] P. Stark, "Conservative statistical post-election audits," *Ann. Appl. Stat.*, vol. 2, pp. 550–581, 2008. [Online]. Available: http://arxiv.org/abs/0807.4005

[17] J. Hall, L. Miratrix, P. Stark, M. Briones, E. Ginnold, F. Oakley, M. Peaden, G. Pellerin, T. Stanionis, and T. Webber, "Implementing risk-limiting post-election audits in California," in *Proc. 2009 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE '09)*. Montreal, Canada: USENIX, August 2009. [Online]. Available: http://www.usenix.org/event/evtwote09/tech/full_papers/hall.pdf

[18] P. Stark, "Efficient post-election audits of multiple contests: 2009 California tests," http://ssrn.com/abstract=1443314, 2009, 2009 Conference on Empirical Legal Studies.

[19] ——, "Super-simple simultaneous single-ballot risk-limiting audits," in *Proceedings of the 2010 Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (EVT/WOTE '10)*. USENIX, 2010. [Online]. Available: http://www.usenix.org/events/evtwote10/tech/full_papers/Stark.pdf

[20] M. Higgins, R. Rivest, and P. Stark, "Sharper p-values for stratified post-election audits," *Statistics, Politics, and Policy*, vol. 2, no. 1, 2011. [Online]. Available: http://www.bepress.com/spp/vol2/iss1/7

[21] M. Lindeman and P. B. Stark, "A gentle introduction to risk-limiting audits," *IEEE Security and Privacy*, 2012, to appear. [Online]. Available: http://statistics.berkeley.edu/~stark/Preprints/gentle12.pdf

[22] T. Magrino, R. Rivest, E. Shen, and D. Wagner, "Computing the margin of victory in IRV elections," in *Proceedings of the 2011 Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (EVT/WOTE '11)*. USENIX, August 2011.

[23] D. Cary, "Estimating the margin of victory for instant-runoff voting," in *Proceedings of the 2011 Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (EVT/WOTE '11)*. USENIX, August 2011.

[24] California Secretary of State, "AB 2023 (Saldaña), chapter 122, statutes of 2010 post-election risk-limiting audit pilot program march 1, 2012, report to the legislature," 2012, retrieved 17 March 2012. [Online]. Available: http://www.sos.ca.gov/voting-systems/oversight/risk-pilot/report-to-legislature-3-1-12.pdf

[25] ——, "DRAFT Post-Election Risk-Limiting Audit Pilot Program 2011-2012: Step-by-Step Instructions for Conducting Risk-Limiting Audits," 2012, retrieved 12 March 2012. [Online]. Available: http://www.sos.ca.gov/voting-systems/oversight/risk-pilot/draft-audit-instructions.pdf

Philip B. Stark is Professor of Statistics, University of California, Berkeley. He served on the 2007 California Post Election Audit Standards Working Group and designed and conducted the first risk-limiting post election audits. He is working with the California and Colorado Secretaries of State on pilot risk-limiting audits. For a more complete biography, see http://statistics.berkeley.edu/~stark/bio.pdf.

David A. Wagner is Professor of Electrical Engineering and Computer Science, University of California, Berkeley. He helped lead the 2007 California Top-to-Bottom Review of voting systems and he serves on the Election Assistance Commission's Technical Guidance Development Committee, a federal advisory board charged with helping to draft future voting standards.